

## BAB 2

### PROSEDUR KERJA

#### 2.1 DESKRIPSI PENUGASAN KERJA

Foundation of AI and Life Skills for Gen-Z adalah program pelatihan *Artificial Intelligence* secara online untuk pelajar yang bertujuan bukan hanya untuk memperkenalkan teknologi AI ke pelajar, tapi juga untuk memungkinkan mereka bisa mengangkat perangkat AI, sehingga bisa membuat sesuatu produk yang menciptakan dampak sosial. Berfokus pada komponen utama AI seperti *Data science* (DS), *Natural Language Processing* (NLP) dan *Computer Vision* (CV). Berikut adalah deskripsi pekerjaan sebelum pengerjaan PA:

- a. Mengikuti *pre-test*.
- b. Mengikuti kelas sesi pagi pada pukul 08.00 hingga 11.30 WIB.
- c. Mengikuti kelas sesi siang pada pukul 13.00 hingga 16.30 WIB.
- d. Mengulang materi yang telah disampaikan di kelas sesi pagi dan siang, setelah kelas sesi siang, selama 1 jam (*self-study*).
- e. Mengerjakan latihan individu atau kelompok yang diberikan oleh *homeroom* atau *domain coach* saat kelas berlangsung.
- f. Mengerjakan tugas yang diberikan *homeroom* atau *domain coach* hingga batas waktu tertentu.
- g. Mengerjakan *mini project* yang diberikan *homeroom* atau *domain coach* hingga batas waktu tertentu
- h. Mengikuti *post-test*.

Selama pengerjaan PA memiliki peran *problem scoping*, membuat jurnal, PPT, dan laporan akhir, dengan deskripsi pekerjaan sebagai berikut:

- a. Membuat *problem scoping* dalam projek akhir mengenai Analisis Bank Marketing Menggunakan Klasifikasi Machine Learning Untuk Memperoleh Segmentasi Klien Membuka Akun Rekening Deposito Berjangka.
- b. Menyusun PPT untuk presentasi kelompok.

Selain keterampilan AI ini, memiliki 6 fokus pembelajaran.

### **2.1.1 Logika dan Konsep Teknologi AI**

Pada pembelajaran mengenai logika dan konsep teknologi AI peserta akan mengikuti pembelajaran mengenai pengembangan teknologi mulai dari sejarah dari *Artificial Intelligence* (AI) dan prinsip-prinsip dasar AI serta mempelajari metode-metode yang ada pada *domain* AI. Diharapkan dari pembelajaran ini peserta memiliki wawasan dan perkembangan tentang AI. Selain itu, penilaian yang dilakukan melalui dokumen hasil dari kegiatan pada detail pembelajaran, jawaban ujian, dan materi secara daring.

### **2.1.2 Metode Penelitian AI**

Pada pembelajaran mengenai metode penelitian AI peserta akan mengikuti pembelajaran mengenai konsep-konsep analisis dan statistis dalam *domain* AI, yaitu *Data science* (DS), *Computer Vision* (CV), dan *Natural Language Proccessing* (NLP). Diharapkan dari pembelajaran ini peserta mampu menggunakan analisis dan statistis yang terdapat pada *domain* AI untuk melakukan pemodelan AI dan analisis masalah. Selain itu, penilaian yang dilakukan melalui dokumen hasil dari kegiatan pada detail pembelajaran, jawaban ujian, dan materi secara daring.

### **2.1.3 Siklus Proyek AI**

Pada pembelajaran mengenai siklus proyek AI peserta akan mengikuti pembelajaran mengenai framework sebagai kerangka dalam membuat aplikasi AI. *AI cycle project* terdiri dari tahap *Data Scoping*, *Data Acquisition*, *Data Exploration*, *Modelling*, dan *Evaluation*. Diharapkan dari pembelajaran ini peserta mampu merancang dan mengimplementasikan *AI cycle project* dengan metode-metode yang ada pada *domain* AI dalam kehidupan sehari-hari. Selain itu, penilaian yang dilakukan melalui dokumen hasil dari kegiatan pada detail pembelajaran, jawaban ujian, dan materi secara daring.

### **2.1.4 Pemrograman Python**

Pada pembelajaran mengenai pemrograman *python* peserta akan mengikuti pembelajaran tentang *python basic* dan *advanced* yang mana dimulai dari tahap instalasi *python* program, tata tulis dalam pemrograman *python*, logika, dan tipe

data operator sampai dengan latihan-latihan yang mengacu pada *domain* AI. Diharapkan dari pembelajaran ini peserta mampu menggunakan Bahasa *python* programming untuk menyelesaikan masalah sederhana berbasis AI dan membuat aplikasi AI. Selain itu, penilaian yang dilakukan melalui dokumen hasil dari kegiatan pada detail pembelajaran, jawaban ujian, dan materi secara daring.

### **2.1.5 Etika Profesi dan Kewirausahaan**

Pada pembelajaran mengenai etika profesi dan kewirausahaan peserta akan mengikuti pembelajaran berupa materi keterampilan interpersonal skill, etika kerja, manajemen waktu, komunikasi bisnis, pola berpikir CV dan teknik wawancara. Diharapkan dari pembelajaran ini peserta mampu menggunakan softskills dan hardskills dalam dunia industry dan perusahaan. Selain itu, penilaian yang dilakukan melalui dokumen hasil dari kegiatan pada detail pembelajaran, jawaban ujian, dan materi secara daring.

### **2.1.6 Projek Akhir**

Pada pembelajaran mengenai projek akhir mengikuti pembelajaran tentang materi berupa tahapan-tahapan didalam memodelkan system *Artificial Intelligence* dalam bentuk produk aplikasi. Diharapkan dari pembelajaran ini peserta mampu produk aplikasi *Artificial Intelligence* sebagai representasi bisnis dan solusi penyelesaian masalah berbasis AI. Selain itu, penilaian dilakukan melalui dokumen hasil dari kegiatan pada detail pembelajaran, jawaban ujian dan materi secara online. Sedangkan untuk penilaian atas pengerjaan projek akhir yang dilakukan secara individu atau berkelompok dilakukan penilaian dalam bentuk dokumen yang telah ditetapkan dan disediakan oleh mitra sebagai standar penilaian dan pelaksanaan projek akhir.

## **2.2 DASAR TEORI**

### **2.2.1 Google Colaboratory**

*Google Colaboratory* (Colab) adalah *tools* baru yang dikeluarkan oleh yang dibuat oleh Google Internal Research. *Tools* ini membantu para *Researcher* dalam mengolah data untuk keperluan belajar pada pengolahan data menggunakan *Machine Learning* [1]. Google Colab sendiri merupakan bentuk modifikasi dari

Jupyter Notebook yang disediakan untuk Google, dimana platform ini sering dimanfaatkan untuk *Machine Learning* dan *Deep Learning* [2]. Penggunaan *tools* ini mirip dengan Jupyter Notebook. Google Colab memiliki penyimpanan berupa Google Drive dimana *tools* ini berjalan dengan sistem *Cloud*. Google Colab ini menyediakan layanan GPU gratis kepada penggunanya sebagai backend komputasi dan dapat digunakan selama 12 jam pada suatu waktu [1].

### 2.2.2 *Data science*

*Data based science* atau *data science* adalah sebuah studi Ilmu *Data Based* atau Ilmu Data adalah sebuah studi interdisipliner yang mengeksplorasi metode ilmiah dan cara mengekstrak pengetahuan atau wawasan tentang data dari berbagai bentuk, tidak hanya terstruktur tetapi juga tidak terstruktur. Tahapan dalam *data science* mirip dengan yang ada dalam 3 tahap Penemuan Pengetahuan dalam Basis Data (KDD) :

#### 1. *Pre-processing data*

Untuk menganalisis kumpulan data multivariat sebelum penambahan data. Pre-processing juga bertujuan untuk membersihkan data dengan menghilangkannya dari pengamatan yang kekurangan noise atau data.

#### 2. Data Mining umumnya melibatkan enam kelompok tugas yaitu :

- Deteksi anomali (deteksi deviasi)
- Identifikasi catatan data yang tidak biasa, data menarik atau kesalahan data yang memerlukan penyelidikan lebih lanjut.
- Pembelajaran aturan asosiasi (ketergantungan pemodelan)
- Menemukan hubungan antar variable
- *Clustering* adalah tugas menemukan kelompok dan struktur dalam data yang dalam berbagai cara.
- Klasifikasi adalah tugas generalisasi struktur yang diketahui untuk diterapkan pada data baru.
- Regresi mencoba menemukan fungsi yang memodelkan data dengan sedikit kesalahan.
- Peringkasan memberikan representasi data yang solid, termasuk visualisasi dan laporan.

3. Validasi hasil disusun untuk memahami dan memberi hasil dari jenis prediksi yang sama atau untuk menemukan jenis pola umum yang sama, terlepas dari data eksperimen yang diberikan [3].

### 2.2.3 *Machine Learning*

*Machine Learning* (ML) adalah pendekatan AI yang banyak digunakan untuk menggantikan atau meniru perilaku manusia untuk memecahkan masalah atau mewujudkan otomatisasi. Seperti namanya, ML mencoba meniru cara manusia atau makhluk cerdas belajar dan menggeneralisasi proses. Setidaknya ada dua aplikasi utama dalam ML, klasifikasi dan prediksi. Ciri khas ML adalah adanya proses pelatihan, pembelajaran atau pelatihan. Oleh karena itu, ML membutuhkan data pembelajaran yang dikenal sebagai data pelatihan. Klasifikasi adalah metode ML yang digunakan oleh mesin untuk mengurutkan atau mengklasifikasikan objek berdasarkan karakteristik tertentu sementara manusia mencoba membedakan objek satu sama lain. Sedangkan prediksi atau regresi digunakan oleh mesin untuk menebak keluaran dari data masukan berdasarkan data yang telah dipelajari selama pelatihan. Metode ML yang paling populer adalah Sistem Pengambilan Keputusan, *Support Vector Machines* (SVM), dan *Neural Network* [4].

#### 1. *Stratified Sampling*

*Stratified random sampling* adalah suatu teknik pengambilan sampel dengan memperhatikan suatu tingkatan (strata) pada elemen populasi. Elemen populasi dibagi menjadi beberapa tingkatan (stratifikasi) berdasarkan karakter yang melekat padanya. Dalam elemen populasi dikelompokkan pada tingkatan-tingkatan tertentu dengan tujuan pengambilan sampel akan merata pada seluruh tingkatan dan sampel mewakili karakter seluruh elemen populasi yang heterogen [5].

Konsep *stratified sampling* penting ketika membangun sebuah model baik untuk klasifikasi maupun regresi. Untuk menghindari *over fitting* pada data, maka harus mengimplementasikan sebuah “*cross validation*” dimana harus dipastikan bahwa fitur-fitur yang memiliki pengaruh besar terhadap label terdistribusi secara merata. *Split* data menjadi data training dan data test dan implementasikan sebuah *stratified shuffle split* dari *library Python*. Model klasifikasi yang digunakan adalah:

##### a. *Logistic Regression*

Pada *python* terdapat dalam *library* *scikit\_linear* dengan bernama *Logistic Regression*, fungsi *logistic* secara umum dapat dirumuskan dengan

$$y_i = f(x_i\beta) + \varepsilon_i$$

Mengingat bahwa untuk regresi linear kita perlu menyesuaikan model dengan meminimumkan jumlah *squared errors*, yang mana berakhir dengan memilih yang memaksimumkan *the likelihood of data*.

#### b. *K-Neighbour Classification*

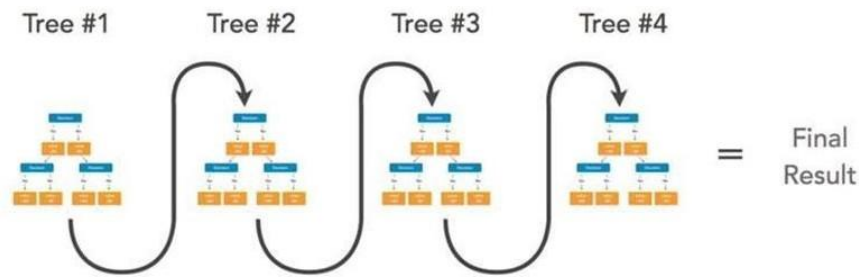
Algoritma KNN atau *K-Nearest Neighbour* merupakan algoritma yang sederhana, algoritma *supervised Machine Learning* yang mudah digunakan untuk menyelesaikan permasalahan klasifikasi dan regresi. Algoritma KNN mengklasifikasikan berdasarkan jarak terdekat dengan objek lainnya (tetangga). Cara kerja algoritma dengan menentukan nilai K atau jumlah tetangga yang akan digunakan untuk pertimbangan penentuan kelas. Selanjutnya penghitungan jarak dari data baru ke masing-masing data point di *dataset*. Dengan sejumlah K data dengan jarak terdekat, dapat ditentukan kelas dari data baru tersebut. Algoritma *K-Neighbour Classification* menggunakan kemiripan antara data baru dengan sejumlah data yang terdekat yang telah tersedia. KNN bekerja dengan menyimpan set data training dan belajar learning pada saat membuat prediksi secara *real time* [6].

#### c. *Linear Support Vector Machine (SVM)*

Algoritma SVM merupakan salah satu metode klasifikasi *supervised learning* yang bekerja dengan pemetaan *non linear* untuk mengubah data pelatihan asli ke dimensi yang lebih tinggi [7].

#### d. *Gradient Boosting Classifiers*

Model *Gradient Boosting Classifiers* merupakan algoritma *Machine Learning* yang menggunakan *ansambel* dari *Decision Tree* untuk memprediksi nilai. Menggunakan beberapa model *Machine Learning* secara bersama untuk menghasilkan prediksi yang tepat. Pendekatan dengan menyambungkan hasil akhir dari *Decision Tree* dan perbaikan terhadap *error*nya merupakan algoritma dari *Gradient Boosting Classifiers* [8].



Gambar 3.1 Algoritma GBC [8]

Tujuan model ini berkaitan dengan membuat model hubungan :

- Hubungan saldo dengan status pernikahan (“*balance*”- “*marital*”)
- Grup yang memiliki pinjaman pribadi dan grup status pernikahan-pendidikan (“*loan*”- “*marital education group*”) [8]

e. *Decision Tree*

Seperti SVM, *Decision Tree* adalah algoritma *Machine Learning* seba guna yang dapat bekerja keduanya untuk tugas klasifikasi maupun regresi dan bahkan tugas dnegan banyak *output*. *Decision Tree* komponen dasar pada model *Random Forests*. *Decision Tree* menggunakan struktur pohon untuk merepresentasikan jumlah kemungkinan jalur-jalur keputusan dan hasil / keluaran untuk masingmasing jalur. Pohon yang dibangun akan mengandung simpul keputusan (*decision nodes*) (yang mana menanyakan sebuah pertanyaan dan mengarahkan dengan berbeda tergantung dengan jawaban dan simpul daun (*leaf nodes*) [9].

f. *Random Forest Classifier*

*Random forest* adalah algoritma ensemble learning yang dibangun dari pohon keputusan. Algoritma ini membuat beberapa pohon keputusan menggunakan data *boot strap* (metode resampling yang menggunakan algoritma random sampling yang diganti beberapa kali) dan secara acak memilih *subset* variabel di setiap pohon keputusan.

g. *Artificial Neural Network*

*Artificial Neural Network* (ANN) atau Jaringan Syaraf Tiruan merupakan suatu sistem pemrosesan informasi dengan suatu karakteristik menyerupai sistem saraf pada manusia yang dapat memecahkan masalah SVM dan KNN dengan melakukan training data yang besar dan ANN memiliki kemamuan untuk mentoleransi atau

diagnosis kesalahan sehingga dapat menghasilkan prediksi yang baik. Selain itu, metode ini juga dapat digunakan untuk memodelkan hubungan yang kompleks antara masukan (*input*) dan keluaran (*output*) dalam menemukan pola-pola pada data [4].

#### *h. Naive Bayes Classifier*

*Naive bayes classifier* merupakan metode klasifikasi berakar yang menggunakan metode probabilitas dan statistik untuk memprediksi suatu peluang terhadap suatu kondisi, dan salah satu kegunaannya ialah untuk klasifikasi dokumen teks. Metode yang juga dikenal sebagai *Gaussian NB* ini menerapkan teknik *supervised* klasifikasi objek di masa depan dengan menetapkan label kelas ke instance/catatan menggunakan probabilitas bersyarat. *Distribusi Gaussian* adalah asumsi pendistribusian nilai kontinu yang terkait dengan setiap fitur berisi nilai numerik. Ketika diplot, akan muncul kurva berbentuk lonceng yang simetris tentang rata-rata nilai fitur [10].